Folding pathways of a helix-turn-helix model protein

Daniel Hoffmann* and Ernst-Walter Knapp
Freie Universität Berlin,
Fachbereich Chemie, Institut für Kristallographie,
Takustr. 6, D-14195 Berlin, Germany

J. Phys. Chem., in press

Abstract

A small model polypeptide represented in atomic detail is folded using Monte Carlo dynamics. The polypeptide is designed to have a native conformation similar to the central part of the helix-turn-helix protein ROP. Starting from a β-strand conformation or two different loop conformations of the protein glutamine synthetase, six trajectories are generated using the so-called window move in dihedral angle space. This move changes conformations locally and leads to realistic, quasi-continuously evolving trajectories. Four of the six trajectories end in stable native-like conformations. Their folding pathways show a fast initial development of a helix-bend-helix motif, followed by a dynamic behaviour predicted by the diffusion-collision model of Karplus and Weaver. The phenomenology of the pathways is consistent with experimental results.

*Present address: German National Research Center for Information Technology, GMD-SCAI, Schloss Birlinghoven, D-53754 Sankt Augustin, Germany; e-mail daniel.hoffmann@gmd.de; URL http://www.gmd.de/SCAI/people/hoffmann.html

Introduction

One of the most intriguing problems in molecular biology is the decryption of the protein folding code. There is a wealth of experimental results providing insights into kinetics and thermodynamics of the folding process [1, 2]. They point to a delicate interplay of hydrophobic and electrostatic interactions which guide the formation of secondary and tertiary structures. Computer experiments on protein folding using Monte Carlo (MC) methods with simplified lattice models also contributed very much to our understanding of the principles of protein folding [3, 4]. With these simplified protein models general aspects of protein folding can be studied. However they are not designed to reflect the behavior of a particular existing protein at the atomic level of description. The latter is the domain of conventional molecular dynamics (MD) [5, 6, 7, 8]. Unfortunately conventional MD cannot be used to address the protein folding problem for the time being. The reason for this becomes clear if we consider that the the typical time propagation step in MD is 1 fs, and that the CPU-time per evaluation of the energy function is of the order of 1 s. Hence typical folding processes lasting 1 ms to 1 s would require CPU-times of 10^{12} s to 10^{15} s.

Off-lattice MC dynamics is an alternative that can extend the time range accessible to simulation into the folding regime [9]. We have proposed a MC method [10, 11, 12] which combines a detailed protein model, with dihedral angles as continuous degrees of freedom, and an efficient algorithm for the generation of new conformations. This so-called window algorithm simulates the evolution of the polypeptide conformation by series of local conformational changes, each one restricted to a window, i. e. a randomly selected short stretch of polypeptide backbone. As has been shown earlier [12] the local move, with its cooperative changes of dihedral angles in the window, performs far better than a simple MC move where torsion angles of the polypeptide backbone are changed independently and thus global conformational changes are generated. Furthermore window moves simulate the dynamical time evolution at least two orders of magnitude faster than MD [13, 12].

In this paper we study MC trajectories for a small model protein in detail. These trajectories provide an atomistic view of possible mechanisms in the protein folding process and offer interpretations for experimental results like the fast formation of secondary structure [14, 15], the existence of transient non-native conformations [16, 17, 18], or the existence of multiple pathways [19]. It turns out that the time evolution observed in the trajectories is in good agreement with predictions made by the diffusion-collision model (DCM) of Karplus and Weaver [20]. This model allows a quantitative description of the folding mechanism first proposed by Ptitsyn and Rashin [21], and later supported by Kim and Baldwin [22], who coined the name "framework model". The DCM essentially states that the first step

in folding is the formation of microdomains, e. g. α-helices, which diffuse relatively to each other and eventually collide and coalesce with a certain probability. In this way new and larger domains are formed, which again collide and coalesce, etc. In this sense the formation of an α-helical hairpin can be understood as an elementary event of the DCM.

The model system

The simulation of protein folding with a detailed model demands great amounts of CPU-time. Therefore it is important to choose a protein that is as simple as possible. It nevertheless should have features typical of proteins, i. e. a native state with secondary and tertiary structure. These structural elements are stabilized by hydrogen bonds and hydrophobic interactions. Hence, the model should consider corresponding energy terms. ROP is a simple protein that has secondary structure and tertiary contacts [23]. It forms an α-helix-turn-α-helix, that is a so-called α-helical hairpin motif. The central part of ROP, consisting of the 26 residues from 18 to 43, is used as a template for the construction of a model polypeptide. Three types of amino acids are used. The five residues in the central turn are replaced by glycines (G) which are known to have a high propensity to form turns. The residues responsible for interhelical contacts are assigned to residues of type X that differ from alanines only in the increased attraction between C_β-atoms of X residues, mimicking a hydrophobic interaction. This attraction is modeled by a Lennard-Jones potential with a well-depth of 2.0 kcal/mol, a value that is motivated by the free energy changes for transfer of typical hydrophobic amino acids from a non-polar to a polar solvent [24, 25]. All remaining residues are replaced by alanines (A). Alanines are often found in α-helical secondary structure of proteins, and in MD simulations of polyalanine α-helices are formed *in vacuo* [26, 27]. The whole sequence of the model polypeptide reads AXAAXAAAXXGGGGGXXAAAXAAAXA. The terminal alanines are blocked with neutral acetamide and N-methyl-amide groups to avoid strong Coulombic interactions. Since each of the sidechains of A and X is represented by a single, so-called extended C_β-atom [5], there are no torsional degrees of freedom in the sidechains. The similarity of the model polypeptide with the original ROP is merely a structural one, insofar as it has a high propensity to form a α-helical hairpin. With respect to other properties there may be significant differences between model polypeptide and ROP, e.g. the ROP monomer is not stable in aqueous solution [28].

The bond lengths and bond angles of the model polypeptide, as well as the dihedral angles $\omega_i(C_{\alpha,i}C_iN_{i+1}C_{\alpha,i+1})$ are fixed to equilibrium values provided by the parameter set of the MD programme CHARMM22 [5]. Thus the only remaining degrees of freedom of the model polypeptide are the dihedral angles $\phi(C_{i-1}N_iC_{\alpha,i}C_i)$ and $\psi(N_iC_{\alpha,i}C_iN_{i+1})$ in the backbone. The force field is adopted from the MD programme CHARMM, with specific changes to compensate the greater rigidity of the polypeptide model due to the fixed bond lengths and bond angles. This compensation is achieved by replacing the explicit atom pair interactions between sequentially neighbouring amide planes by an effective two-dimensional (ϕ, ψ) torsion potential which implicitely considers the flexibility of the rigidified degrees of freedom. This torsion potential is obtained once by constrained energy minimization

of dipeptides with fixed values of ϕ and ψ but all other degrees of freedom unconstrained, as described by Brooks *et al.* in Appendix 2 of Ref. [5]. Apart from the replacement of the respective atom pair interactions by the torsion potential, the energy function is that of CHARMM [5].

For each move a window is placed randomly on the polypeptide, the conformation is changed in the window, and the energy of the new conformation is evaluated. The difference of energy between the new and the preceding conformation is then used in the criterion of Metropolis *et al.* [29] to decide whether the MC move is accepted or rejected. Conformations generated by this procedure represent a canonical ensemble at the given temperature T. In the present case only windows containing three peptide planes are used because there the acceptance probability is particularly favorable with values ranging from 0.3 to 0.4.

In a first step the rigidified backbone of the model polypeptide was fitted to the considered central part of the x-ray structure of ROP by minimizing the root mean square deviation (RMSD) with respect to the backbone atoms. In this ROP-fitted conformation the model polypeptide possesses five interhelical X–X pairs with C_β-C_β distances of less than 6 Å and a well developed system of α-helical hydrogen bonds in each of the two helices. There are no strains in this conformation as can be concluded from energy minimization, that is Metropolis MC at $T = 0$ K, where the conformation changes by less than 1 Å RMSD and the energy drops from -1140 kcal/mol to -1200 kcal/mol. Thus the designed amino acid sequence has a native conformation close to that of the ROP-fitted conformation. This assumption is supported further by eight simulated annealing simulations with initial temperatures of 1000–3000 K starting from the ROP-fitted conformation of the model polypeptide. In these simulations the conformation of lowest energy (-1215 kcal/mol), the "reference structure", had a RMSD of 1.35 Å to the ROP-fitted conformation.

It has been shown elsewhere [12] that window moves are able to simulate the folding of this model polypeptide. Here we analyse the folding process in more detail. Six trajectories were generated at $T = 450$ K, a temperature low enough for the designed native conformation being stable, and high enough to facilitate folding within a reasonable short amount of CPU-time. The elevated temperature can compensate for the missing aqueous solvent, whose presence would weaken intra-polypeptide interactions like hydrogen bonding. Furthermore isomerization barriers are still somewhat to high compared with those encountered in MD simulations despite the adaptation of the energy function to the rigidified polypeptide model. Thus $T = 450$ K corresponds to a lower temperature for MD simulations in aqueous solution.

Results and discussion

Six trajectories were generated. Four of them (β-trajectories) started from a β-strand conformation ($\phi = -120°, \psi = 120°$). This conformation is quite elongated, and thus far away from the "native" helical hairpin structure, but unlike the fully extended conformation, for the β-strand the interactions of neighbouring peptide planes are in a local minimum. Hence the β-strand seems to be a reasonable model for a denaturated state. In order to investigate the sensitivity of the results to the initial conditions, the remaining two trajectories (loop-trajectories) started from two other conformations, namely the one of loop Glu 13 – Asn 39 and of loop Lys 163 – Gln 189 (Fig. 1), respectively, of chain F in the protein complex glutamine synthetase (PDB code 2gls) [30]. These two loops have completely different structures and are devoid of helical turns. Hence there is no bias towards the α-helical hairpin. Five out of the six trajectories end up in helix-turn-helix conformations which after minimization have lower energies than the reference structure introduced in the previous section. The two loop-trajectories and two of the β-trajectories are well equilibrated after about 10^6 MC scans of window moves (in a MC scan the window is placed randomly on the polypeptide backbone as many times as there are possible window positions). In the end they show only modest fluctuations of the energy and of structural quantities like the radius of gyration. These four trajectories deliver conformations close to the reference structure (all-atom root mean square deviations 2.3, 2.7, 3.2, 3.8 Å, respectively), and the number of interhelical X–X pairs with C_β-C_β distance smaller than 6 Å is five to six, which is identical to that of the reference structure. The mean energies in the equilibrated parts of the four trajectories lie at about -1190 kcal/mol. The other two β-trajectories lead to a single long helix, and a helix-turn-helix motif with some left handed helical turns, respectively. The single helix shows fraying at the termini and also bending dynamics, but develops no stable loop between two helices. Due to the missing X–X interactions the mean energy (-1175 kcal/mol) of the trajectory is significantly higher than of the other trajectories. The same holds true for the structure with the left handed helix turns, which prevent the formation of all possible interhelical X–X pairs. These two trajectories hence represent different free energy minima above the minimum corresponding to the "native" helical hairpin. They could reach this lower minimum by breaking some helical hydrogen bonds in the case of the long helix, or by expansion and refolding with the correct righthanded helical turns in the case of the other trajectory. Both trajectories explore conformations towards these barriers but do not cross them during the simulation of $2 \cdot 10^6$ MC scans. In the following we focus mainly on the folding paths observed in the two equilibrated β-trajectories, which we call trajectory (1) and (2), respectively. These two trajectories are representative for those four trajectories which converge into α-helical hair-

pins. If not mentioned otherwise the described observations refer to both trajectories.

Within the first few tens of thousands of MC scans the elongated β-strand relaxes into a mainly α-helical conformation (Fig. 4(b) and Fig. 4(c)). The initiations of the helices take place almost simultaneously at different locations of the polypeptide, but preferentially near the termini. The only exception is the C-terminal helix in trajectory (1) which begins to grow at residues 17, 18 and 19. Each helix grows until it reaches the nearest terminus, or until it is stopped by a multiple turn structure near the center of the polypeptide at one end of the stretch of glycines. At this stage the turn structure consists mainly of X and A residues, and most of the glycines are incorporated into one of the two helices. This is particularly noteworthy because usually glycines act as helix breakers. However there are experiments indicating that non-native conformations can have appreciably populated non-native secondary structures [17, 18]. In this context it is interesting to note that peptide fragments of myohemerythrin [16] and plastocyanin [31] which include the loop and turn regions of the native proteins clearly show α-helix content.

The turn region functions as a buffer between the N- and C-terminal helices, preventing them from unification. The helix growth is accompanied by a sharp drop in energy (Figs. 2(a) and 3(a)) from -1008 kcal/mol for the initial β-strand conformation to about -1175 kcal/mol, mainly originating from the formation of helical hydrogen bonds and other attractive sequence local interactions. The increasing helix content is also visible in the contraction from 25 Å to 10 Å radius of gyration R_{gyr}, or 82 Å to 32 Å end-end-distance R_{ends} (Figs. 2(c) and 3(c)). The observed fast formation of secondary structure as a first step in the folding process is consistent with experimental findings for a variety of proteins, like cytochrome c and β-lactoglobulin [14], ribonuclease A [15], lysozyme [19], or *Escherichia coli* trp aporepressor [32].

The phase of helix formation ends after 10^5 MC scans and can be clearly distinguished from the following second phase in which the helix content remains essentially constant but R_{ends} fluctuates considerably. In other words the two α-helices are diffusing relatively to each other as quasi rigid entities with only the interhelical angle changing randomly (Figs. 2(d) and 3(d)). In trajectory (1) this diffusion leads to a near coalescence of the two helices after about
$3 \cdot 10^5$ MC scans where R_{ends} and R_{gyr} drop sharply. Simultaneously the number of pairs of non-neighbouring hydrophobic X residues with C_β-C_β distances less than 6 Å increases transiently from one or two to six. At the same time the short N-terminal helix unravels almost completely (Fig. 4(b)). During this near-coalescence the energy trace shows no dramatic change. The new more compact conformation is unstable and decays into a more elongated one, while the short N-terminal helix recovers. The diffusion of

the angle between the two helices lasts for about $3 \cdot 10^5$ and 10^5 MC scans in trajectory (1) and (2), respectively. The shorter diffusion phase in trajectory (2) may account for the lack of near-coalescence events like that observed in trajectory (1). For both trajectories the energy has a value of -1160 ± 10 kcal/mol during the helix angle diffusion.

It is notable that trajectories (1) and (2), although being formally very similar and undistinguishable if for example only the development of the helix content would be monitored, represent two different folding pathways. In (1) the turn forms in the N-terminal half of the sequence (Fig. 4(b)), whereas in (2) it forms in the C-terminal half (Fig. 4(c)). Thus in (1) a short N-terminal helix coalesces with a longer C-terminal one, while in (2) a short C-terminal helix joins a longer N-terminal one (Figs. 5 and 6). The existence of multiple folding pathways has also been suggested to explain experiments on ribonuclease [33] and lysozyme [19].

The period of relative diffusion of the interhelical angle is terminated by the coalescence of the two helices. The coalescence is initiated by the formation of a cluster of hydrophobic X residues at the shorter helix, inducing a sharp reverse turn and bending the shorter helix towards the longer one. Then this bent conformation leads to the formation of the first interhelical hydrophobic X-X pairs. The number of X-X pairs is still small, indicating the non-native character of the conformations in this phase. The non-native character can also be seen from the fact that shortly after coalescence the glycines in the center of the sequence are still part of one of the helices and hence the α-hairpin is quite asymmetric with a shorter and a longer helix (trajectory (1) after 460000 MC scans, Fig. 5, and trajectory (2) after 160000 MC scans, Fig. 6).

Now a new type of movement can be observed. The respective shorter helix begins to creep along the longer one. This movement is driven by the attraction of X residues in interhelical X-X pairs and leads to the formation of an increasing number of X-X pairs. The creeping generates a pull which is transmitted onto the longer helix through the connecting reverse turn. Eventually this pull forces the glycines out of the helical turn into the reverse turn and thus the reverse turn expands at the expense of the longer helix. Finally the lengths of the two helices making up the hairpin are approximately equal in length. During the process of creeping, which lasts for $1.6 \cdot 10^5$ MC scans, the energy drops by 20 kcal/mol to about -1190 kcal/mol. This drop is mainly due to the formation of X-X pairs, rising in number from one to about five. As a further result of the process of creeping, R_{ends} falls from 15 Å, shortly after coalescence, to 4–5 Å (Figs. 2(c) and 3(c)), i. e. the hairpin is complete.

After the creeping process has stopped and a compact helical hairpin conformation is reached, the structural fluctuations of the hairpin are reduced considerably. In the picture of the folding funnel [34] the trajectories have reached a thermodynamic bottleneck, where the multiple folding path-

ways approach the native state from different sides and are slowed down by entropic barriers. Nevertheless further rearrangements can be observed. For example in trajectory (1) after $1.2 \cdot 10^6$ MC scans there is a cooperative reorganization of the two termini which goes along with an increase of helix content by one residue, and an increase in the number of X-X pairs to a value fluctuating between five and six (Fig. 4(b)). Due to these changes the energy goes down from -1190 kcal/mol to about -1200 kcal/mol. In trajectory (2) a remarkable development takes place at 10^6 MC scans (Fig. 4(c)) when the glycines in the turn region, which previously had flickered between helical and other turn conformations, are forming a short α-helix which is stable for a few thousand MC scans. This process is also visible in the energy trace as a transient increase by about 10 kcal/mol. At the same time the number of hydrophobic X-X pairs decreases momentarily by two. The glycine helix collapses very suddenly, and all but one glycine are finally stabilized in a non-helical turn conformation. After the collapse of the glycine helix, energy and number of X-X pairs return to their previous levels, but the conformation has changed to a near-native one with a stable non-helical turn of glycines between two helices.

The folding pathways in trajectories (1) and (2) are consistent with the diffusion-collision model (DCM) of Karplus and Weaver [20] which provides a theoretical framework for protein folding and is supported by a large number of experimental findings [35]. As pointed out in the introduction, the essence of the DCM is that in the folding process microdomains are formed, e. g. α-helices, which diffuse relatively to each other. Eventually they collide and then coalesce with a certain probability. This behaviour is observed in trajectories (1) and (2). The helices are formed and diffuse relatively to each other as quasi rigid bodies by random changes of the interhelical angle. In the DCM, a collision of microdomains does not necessarily imply that they remain together. This has been observed in trajectory (1) where after a first collision the two helices separate again and continue their relative diffusion. The DCM further allows that "after partial collapse and/or weak coalescence of microdomains to a more compact structure with a non-native conformation, the attainment of the native conformation might involve surface diffusion in one or two dimensions" [35]. The creeping motion observed in both trajectories is such a one-dimensional diffusion. It is also reminiscent of the reptation movement introduced by de Gennes [36] for polymers in polymer melts and hence could be called self-reptation. Note that this self-reptation movement is different from other dynamic processes where a net increase of helix content is observed in parallel to a formation of a helix dimer (e.g. as in [37]). During self-reptation of the present α-helical hairpin the total helix content remains approximately constant and helical turns are shifted from the Glycine stretch to the neighbouring amino acids.

The DCM not only pictures the folding process qualitatively but also allows the prediction of measurable quantities. For example it is possible to

estimate the time τ_f for the folding of the two helices, i. e. the time between their generation and their ultimate coalescence, using Eq. 1 [38, 35]:

$$\tau_f = l^2/D + L\Delta V(1-\beta)/(DA\beta), \tag{1}$$

where l is a length related to the size of the diffusion space, D is the relative diffusion coefficient of the microdomains, L is a length related to folding/unfolding rates and also to the size of the diffusion space, ΔV is the diffusion volume, A is the collision surface area, and β is the relative coalescence probability or sticking probability for a collision event. Following the procedures for the evaluation of these quantities in Refs. [38, 35] and assuming that the microdomains are two helices of twelve and eight residues, respectively, connected by a loop of six residues, we obtain the following values: $l^2 = 664$ Å^2, $D = 0.138$ Å^2/ps, $L = 16.3$ Å, $\Delta V = 2.30 \cdot 10^5$ Å^3, and $A = 2590$ Å^2. In Refs. [38, 35] β is treated as a free parameter. A large number of simulations of the type presented in this section could be used to determine the probability β more accurately. For a very crude first estimate of β we restrict ourselves to the current data. Since there a near-coalescence occurred only once, the value of β should be of the order of one for the model polypeptide. A value of 0.5 seems a reasonable guess for an order of magnitude estimation. Assuming these parameter values for the quantities in Eq. 1 we find $\tau_f = 2.29 \cdot 10^{-8}$ s. This time can be used to estimate the time corresponding to one MC scan. An inspection of the trajectories shows that the helices coalesce after a diffusion period of about $2 \cdot 10^5$ MC scans. Thus one MC scan is approximately equal to 0.1 ps, and hence each of the MC trajectories runs over about 160 ns. Independently and based on a comparison of dynamic MC with MD simulations over long times we have recently estimated the MC scan to be of the same order of magnitude [12], i. e. about two orders of magnitude larger than the time step of conventional MD. Earlier estimates based on comparisons with the Rouse polymer model [39] yielded a value for the time corresponding to one MC scan which was one order of magnitude larger but did not consider contributions of non-bonded interactions [13].

Previously it was thought that because of the long times involved in the diffusion process, simulations at the atomic level could not be carried out long enough to show aspects of the DCM during protein folding. Therefore simulations have been restricted to the microdomain level of resolution using preformed and explicitely stabilized microdomains [40, 38, 35]. These simulations were valuable to explore general questions concerning the DCM in larger proteins, but of course under these simulation conditions it had to be expected that the trajectories would obey the DCM. Other workers have used simplified lattice models [41]; [42] and found no DCM behaviour. Instead in these models "rather, the helices that form native hairpins are constructed on-site, with folding initiating at or near the turn" [41]. It had

been suspected that the disagreement with DCM may be due to the local moves that had been employed in these lattice simulations which would not allow the diffusion of intact microdomains [35]. Our results show that this argument is not valid in this general form, because the trajectories described above clearly show the diffusion of microdomains despite the use of local moves. It seems more likely that the disagreement with DCM is related to the combined use of lattice models and local moves. Interestingly, Rey and Skolnick [43] compared lattice simulations and off-lattice Brownian Dynamics simulations. Whereas in the lattice simulations the folding of an α-helical hairpin was not consistent with a DCM like process, some of the Brownian Dynamics trajectories clearly followed the DCM scheme.

In the case of our model polypeptide the existence of relatively stable helices certainly promotes the folding according to the DCM. But relatively stable and fast folding polyalanine-based helices are not unusual [44, 45], hence for the given sequence of amino acids, which is dominated by alanines, one should expect DCM like folding. Experimentally, for arbitrary sequences mixtures of various folding mechanisms are observed, including global diffusion of larger parts of the respective proteins (see e.g. Ref. [46]).

Conclusions

Trajectories of a small model protein were produced using a dynamic MC method with window moves. These off-lattice MC moves generate efficiently and realistically conformational changes of a polypeptide. A molecular model in atomic detail with only dihedral angle degrees of freedom, and an energy function derived from a conventional MD model were employed. Due to the ability of MC to generate larger conformational changes per move than MD can generate per step of time propagation, the MC method reaches longer time regimes than MD with the same amount of CPU-time.

Starting from an extended β-strand conformation or from loop conformations of the protein glutamine synthetase, four out of six trajectories equilibrated into a native-like α-helical hairpin conformation within about 10^6 MC scans. Three phases of time evolution can be clearly distinguished in these trajectories: 1. fast formation of two α-helices separated by a turn, 2. diffusion of the interhelical angle, 3. coalescence of the two helices followed by a self-reptation into a native-like helix-turn-helix conformation. All three phases are consistent with experiments and in accord with predictions made by the DCM [35].

The comparison of the folding time predicted for the α-helical hairpin by the DCM with the folding times observed in the MC trajectories yielded a value for the time corresponding to one MC scan of the order of 0.1 ps. This value is in agreement with an independent estimate using a comparison of MD and dynamic MC simulations [12].

Further work is necessary to allow a quantitative comparison of such simulations with experimental results. In particular the energy function has to be refined to consider more quantitatively contributions of the solvent and of a greater variety of sidechains.

Acknowledgment

The authors gratefully acknowledge the help of Fredo Sartori with the force field. The CHARMM code has been provided by Prof. Martin Karplus and Molecular Simulations Inc. This work is supported by the European Union contract ERBCHRXCT930112, the Deutsche Forschungsgemeinschaft, project Kn329/1, and by the Fonds der Deutschen Chemischen Industrie.

References

[1] Alan R. Fersht. Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.*, 5:79–84, 1995.

[2] A. D. Miranker and C. M. Dobson. Collapse and cooperativity in protein folding. *Curr. Opin. Struct. Biol.*, 6:31–42, 1996.

[3] K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan. Principles of protein folding – a perspective from simple exact models. *Protein Science*, 4:561–602, 1995.

[4] Martin Karplus and Andrej Šali. Theoretical studies of protein folding and unfolding. *Current Opinion in Structural Biology*, 5:58–73, 1995.

[5] Bernard R. Brooks, Robert E. Bruccoleri, Berry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculation. *J. Comp. Chem.*, 4:187–217, 1983.

[6] M. Karplus. Internal dynamics of proteins. *Methods in Enzymology*, 131:283–307, 1986.

[7] J. Andrew McCammon and Stephen C. Harvey. *Dynamics of proteins and nucleic acids*. Cambridge University Press, Cambridge, 1987.

[8] W. F. van Gunsteren. The role of computer simulation techniques in protein engineering. *Protein Engineering*, 2:5–13, 1988.

[9] Michael Levitt and Arieh Warshel. Computer Simulation of Protein Folding. *Nature*, 253:694–698, 1975.

[10] E. W. Knapp and A. Irgens-Defregger. Off-Lattice Monte Carlo Method with Constraints: Long–Time Dynamics of a Protein Model Without Nonbonded Interactions. *J. Comp. Chem.*, 13:1–11, 1992.

[11] D. Hoffmann and E. W. Knapp. Polypeptide folding with off-lattice Monte Carlo dynamics: the method. *Eur. Biophysics J.*, 24:387–404, 1996.

[12] D. Hoffmann and E. W. Knapp. Protein dynamics with off-lattice Monte Carlo moves. *Phys. Rev. E*, 53:4221–4224, 1996.

[13] E. W. Knapp. Long Time Dynamics of a Polymer with Rigid Monomer Units Relating to a Protein Model: Comparison with the Rouse Model. *J. Comp. Chem.*, 13:793–798, 1992.

[14] K. Kuwajima, H. Yamaya, S. Miwa, S. Sugai, and T. Nagamura. Rapid formation of secondary structure framework in protein folding studied by stopped-flow circular dichroism. *FEBS Lett.*, 221:115–118, 1987.

[15] J. B. Udgaonkar and R. L. Baldwin. Nmr evidence for an early framework intermediate on the folding pathway of ribonuclease a. *Nature*, 335:694–699, 1988.

[16] H. J. Dyson, G. Merutka, J. P. Waltho, R. A. Lerner, and P. E. Wright. Folding of peptide fragments comprising the complete sequence of proteins. models for initiation of protein folding. i. myohemerythrin. *J. Mol. Biol*, 226:795–817, 1992.

[17] Timothy M. Logan, Yves Thériault, and Stephen W. Fesik. Structural Characterization of the FK506 Binding Protein Unfolded in Urea and Guanidine Hydrochloride. *J. Mol. Biol.*, 236:637–648, 1994.

[18] Ouwen Zhang and Julie D. Forman-Kay. Structural Characterization of Folded and Unfolded States of an SH3 Domain in Equilibrium in Aqueous Buffer. *Biochemistry*, 34:6784–6794, 1995.

[19] Sheena E. Radford, Christopher M. Dobson, and Philip A. Evans. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature*, 358:302–307, 1992.

[20] M. Karplus and D. L. Weaver. Protein-folding dynamics. *Nature*, 260:404–406, 1976.

[21] O. B. Ptitsyn and A. A. Rashin. Stagewise mechanism of protein folding. *Dokl. Akad. Nauk. SSSR*, 213:473–475, 1973.

[22] P. S. Kim and R. L. Baldwin. Specific intermediates in the folding reactions of small proteins and the mechanism of folding. *Annu. Rev. Biochem.*, 51:459–489, 1982.

[23] D. W. Banner, M. Kokkinidis, and D. Tsernoglou. Structure of the ColE1 Rop Protein at 1.7 Å Resolution. *J. Mol. Biol.*, 196:657, 1987.

[24] W. Kauzmann. Some factors in the interpretation of protein denaturation. *Advances in Protein Chemistry*, XIV:1–63, 1959.

[25] C. Chothia. Hydrophobic bonding and accessible surface area in proteins. *Nature*, 248:338–339, 1974.

[26] Bernard R. Brooks. Molecular Dynamics for Problems in Structural Biology. *Chemica Scripta*, 29A:165–169, 1989.

[27] Valerie Daggett, Peter A. Kollman, and Irwin D. Kuntz. A Molecular Dynamics Simulation of Polyalanine: An Analysis of Equilibrium Motions and Helix–Coil Transitions. *Biopolymers*, 31:1115–1132, 1991.

[28] C. Steif, P. Weber, H. J. Hinz, J. Flossdorf, G. Cesareni, and M. Kokkinidis. Subunit interactions provide a significant contribution to the stability of the dimeric four-alpha-helical-bundle protein ROP. *Biochemistry*, 32:3867–3876, 1993.

[29] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

[30] M. M. Yamashita, R. J. Almassy, C. A. Janson, D. Cascio, and D. Eisenberg. Refined atomic model of glutamine synthetase at 3.5 angstroms resolution. *J. Biol. Chem.*, 264:17681, 1989.

[31] H. J. Dyson, J. R. Sayre G. Merutka, Hang-Cheol Shin, R. A. Lerner, and P. E. Wright. Folding of peptide fragments comprising the complete sequence of proteins. models for initiation of protein folding. ii. plastocyanin. *J. Mol. Biol*, 226:819–835, 1992.

[32] C. J. Mann and C. R. Matthews. Structure and stability of an early folding intermediate of *escherichia coli* trp aporepressor measured by far-uv stopped-flow circular dichroism and 8-anilino-1-naphtalene sulfonate binding. *Biochemistry*, 32:5282–5290, 1993.

[33] R. L. Baldwin. Pieces of the folding puzzle. *Nature*, 346:409–410, 1990.

[34] Peter G. Wolynes, Jose N. Onuchic, and D. Thirumalai. Navigating the folding routes. *Science*, 267:1619–1620, 1995.

[35] Martin Karplus and David L. Weaver. Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Science*, 3:650–668, 1994.

[36] P. G. de Gennes. Reptation of a polymer chain in the presence of fixed obstacles. *J. Chem. Phys.*, 55:572–579, 1971.

[37] Michal Vieth, Andrzej Kolinski, Charles Brooks, and Jeffrey Skolnick. Prediction of the folding pathways and structure of gcn4 leucine zipper. *J. Mol. Biol.*, 238:361–367, 1994.

[38] D. Bashford, F. E. Cohen, M. Karplus, I. D. Kuntz, and D. L. Weaver. Diffusion-collision model for the folding kinetics of myoglobin. *Proteins Struct. Funct. Genet.*, 4:211–227, 1988.

[39] P. E. Rouse. A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *J. Chem. Phys.*, 21:1272–1280, 1953.

[40] Sangyoub Lee and Martin Karplus. Brownian Dynamics Simulation of Protein Folding: A Study of the Diffusion-Collision Model. *Biopolymers*, 26:481–506, 1987.

[41] Andrzej Sikorski and Jeffrey Skolnick. Dynamic Monte Carlo Simulations of Globular Protein Folding/Unfolding Pathways: II. α–Helical Motifs. *J. Mol. Biol.*, 212:819–836, 1990.

[42] J. Skolnick and A. Kolinski. Simulations of the Folding of a Globular Protein. *Science*, 250:1121–1125, 1990.

[43] A. Rey and J. Skolnick. Comparison of Lattice Monte Carlo Dynamics and Brownian Dynamics Folding Pathways of α–Helical Hairpins. *Chemical Physics*, 158:199–219, 1991.

[44] Susan Marqusee, Virginia H. Robbins, and Robert L. Baldwin. Unusually Stable Helix Formation in Short Alanine–Based Peptides. *Proc. Natl. Acad. Sci. USA*, 86:5286–5290, 1989.

[45] S. Williams, T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff, and R. B. Dyer. Fast events in protein folding: Helix melting and formation in a small peptide. *Biochemistry*, 35:691–697, 1996.

[46] R. M. Ballew, J. Sabelko, and M. Gruebele. Direct observation of fast protein folding: The initial collapse of apomyoglobin. *Proc. Natl. Acad. Sci. USA*, 93:5759–5764, 1996.

[47] P. Kraulis. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, 24:946–950, 1991.

Figure captions

Figure 1: Conformations from the two trajectories starting from two loop conformations of polypeptide chain F of the protein glutamine synthetase (2glsF). White and black spheres are C_α atoms of glycines and hydrophobic X residues, respectively. The figures were generated using Molscript [47]. N-terminus of polypeptide is always the upper end. Upper left: start conformation corresponding to loop 13 to 39 of 2glsF. Lower left: conformation of lowest energy (-1219 kcal/mol) in trajectory starting from conformation in upper left (after $2.8 \cdot 10^6$ MC scans). All-atom root mean square deviation (RMSD) to reference structure is 3.2 Å. Upper right: start conformation corresponding to loop 163-189 of 2glsF. Lower right: conformation of lowest energy (-1212 kcal/mol) in trajectory starting from conformation in upper right (after $1.6 \cdot 10^6$ MC scans). RMSD to reference structure is 2.7 Å.

Figure 2: Traces of energy, number of pairs of hydrophobic X residues, end-end distance R_{ends}, and interhelical angle in MC trajectory (1). The abscissa of part (d) gives the time in units of 10^5 MC scans. It is used for all four parts of this figur. The value of the respective quantities after every 5000th MC scan is depicted. (a) Conformational energy. (b) Number of pairs of X residues. Two X residues are considered a pair if they are not neighbours in sequence and their C_β-atoms are closer than 6 Å. (c) Distance R_{ends} between first and last C_α-atom of the polypeptide chain. (d) Interhelical angle defined as angle between sum of $C = O$ bondvectors in the first helix (residues 2 to 6), and sum of $O = C$ bondvectors in the second helix (residues 15 to 22).

Figure 3: Trajectory (2). Axes are the same as in Fig. 2. Note however that in part (d) the interhelical angle refers to the angle between the sum of $C = O$ bondvectors of residues 4 to 11, and sum of $O = C$ bondvectors of residues 19 to 24.

Figure 4: Time evolution of (ϕ, ψ) values of each of the 26 residues (structure-dynamograms). (a) Graylevel code for structure-dynamograms (parts (b) and (c)). Each graylevel codes for a rectangular region in the Ramachandran plot around a secondary structure, e. g. white codes for α_R-helical residues. Residues having no canonical structure are coded in black (here not shown for technical reasons). Parts (b) and (c) are structure-dynamograms of MC trajectories (1) and (2), respectively. Conformations are depicted after every 10^4th MC scan.

Figure 5: Coalescence and self reptation in MC trajectory (1). The time is given in numbers of MC scans. C_α-atoms of glycines and hydrophobic X residues are shown as white and black spheres, respectively. Wide ribbons are α-helix turns. N and C indicate N- and C-terminus, respectively. The pictures where generated using Molscript [47].

Figure 6: Coalescence and self reptation in trajectory (2). See also caption of Fig. 5.

Fig. 1 of Hoffmann and Knapp

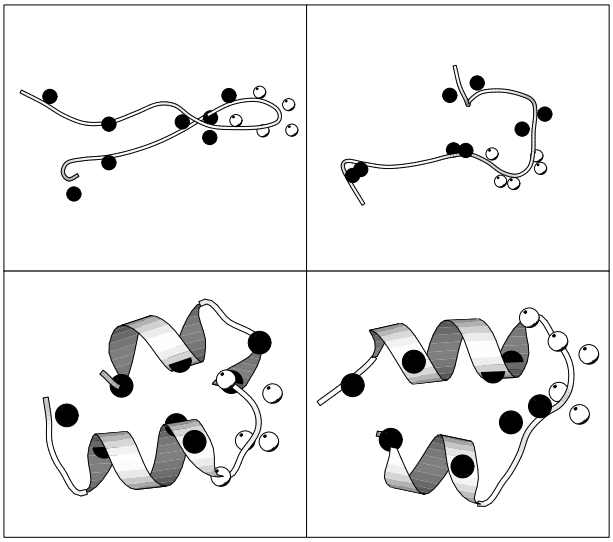

Fig. 2 of Hoffmann and Knapp

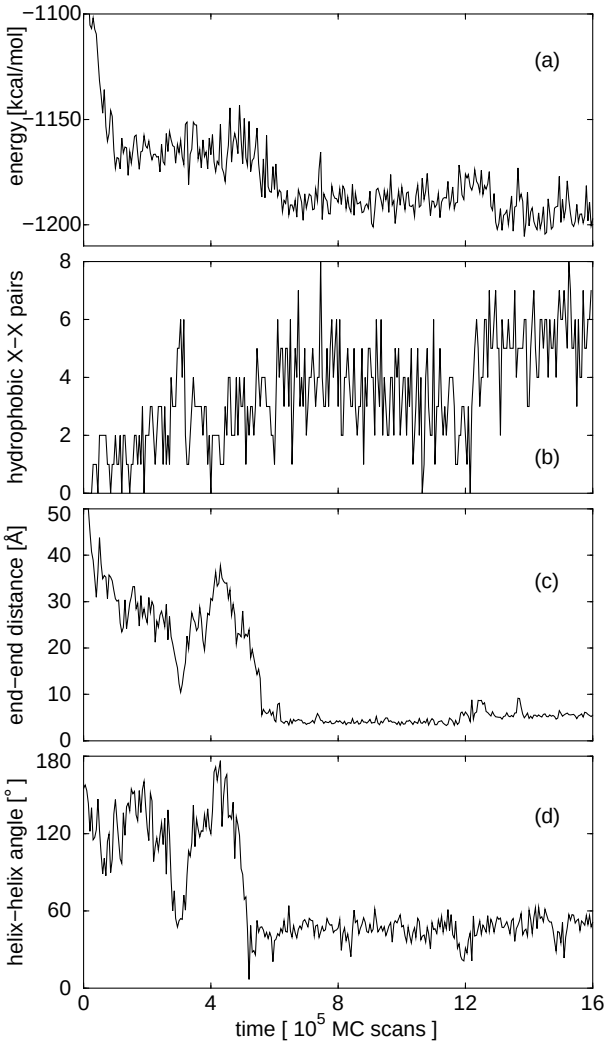

Fig. 3 of Hoffmann and Knapp

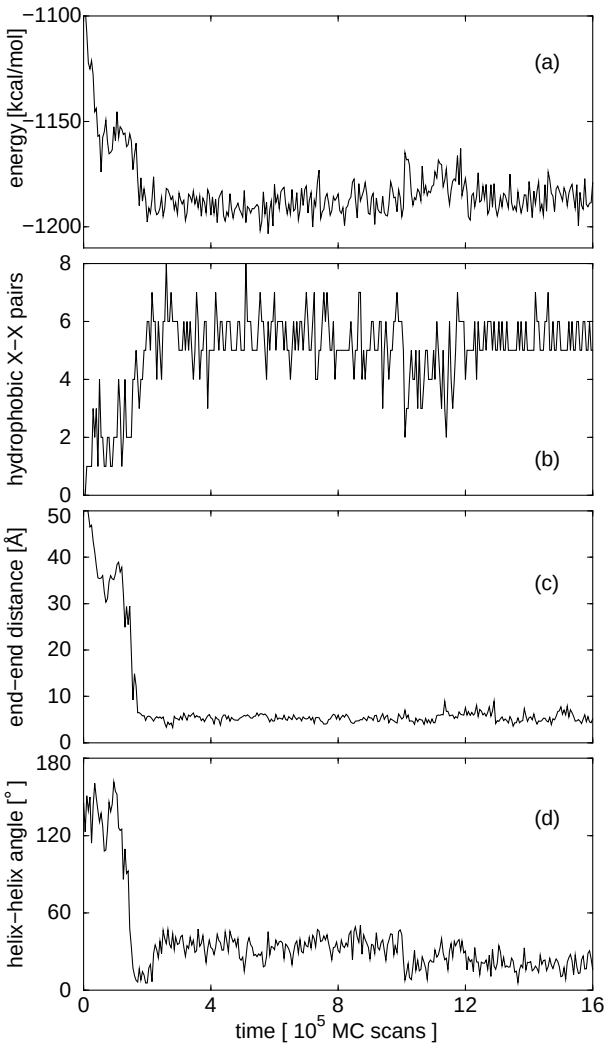

Fig. 4 of Hoffmann and Knapp

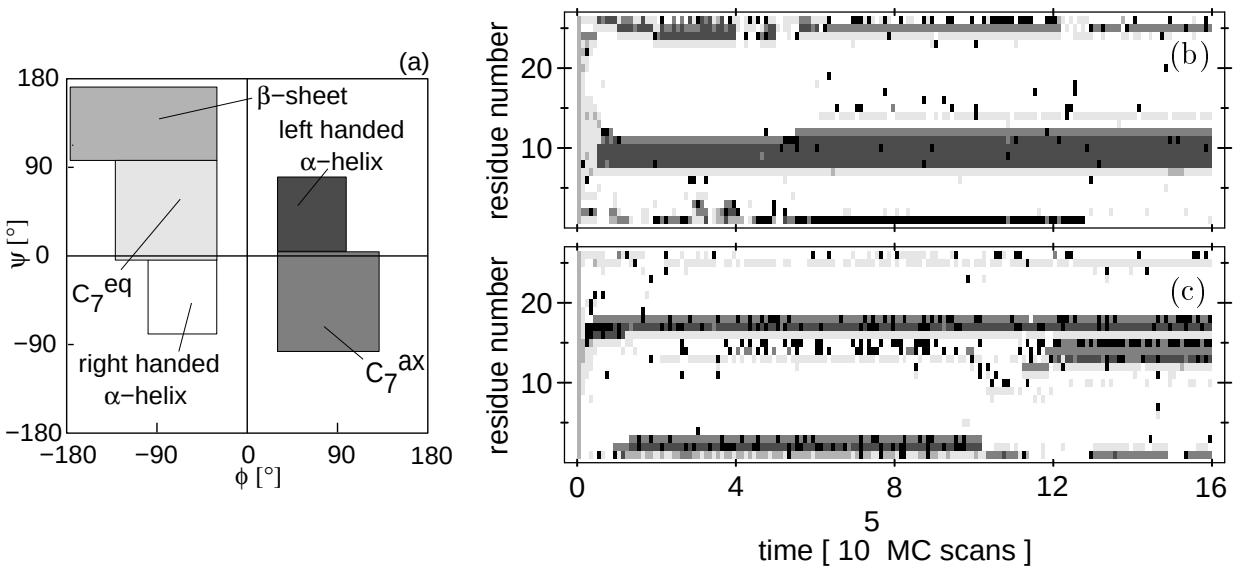

Fig. 5 of Hoffmann and Knapp

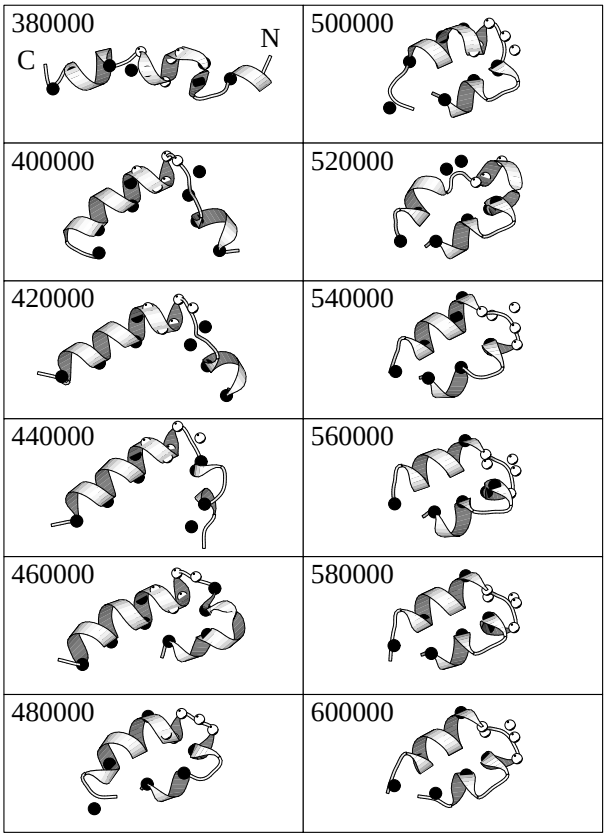

Fig. 6 of Hoffmann and Knapp